1

Foundations of quantum statistical mechanics

1.1 The density operator and probability

Statistical mechanics is concerned with the construction of methods for computing
the expected value of observables important for characterizing the properties of
physical systems, generally containing many degrees of freedom. Starting with a
formally complete detailed description for these many degrees of freedom, proba-
bility theory is used to obtain effective procedures. Quantum statistical mechanics
makes use of two types of probability theory. One of these is the set of natural
probabilities associated with the quantum theory which emerges from its structure
as a Hilbert space. For example, the Born probability is associated with the square
of a wave function. The second is the essentially classical probability associated
with an ensemble of separate systems, each with an a priori probability assigned
by the frequency of occurrence in the ensemble. The quantity which describes both
types of probability in an efficient, convenient way is the density operator.

As an example which illustrates many of the basic ideas, consider a beam of
particles with spin \( \frac{1}{2} \). We shall repeat the resulting definitions later in complete
generality.

The spin states of these particles are represented by two-dimensional spinors
which we denote by the Dirac kets \(|\sigma_z\rangle\) for \( \sigma_z = \pm 1 \), corresponding to the \( z \)
component of the spin \( \sigma \) of the particle. If we perform a filtering measurement to
select a particle of spin \( \sigma' \) with spin \( \sigma'_z = \pm 1 \) in the \( z \) direction, the outcome of the
measurement on a beam of particles with spin \( \sigma_z \) is

\[
|\langle \sigma'_z | \sigma_z \rangle|^2 = \delta_{\sigma'_z, \sigma_z}.
\]

This result can be written as

\[
|\langle \sigma'_z | \sigma_z \rangle|^2 = \text{Tr} P_z (\sigma') P_z (\sigma),
\]
where the projection operator $P_z(\sigma) = |\sigma_z\rangle\langle\sigma_z|$ represents the state of the beam of particles with spin $\sigma$ of definite value $\sigma_z$, and the projection operator $P_z(\sigma')$ represents the experimental question of which value, $\pm 1$, this set of particles has.

If we measure instead a different component of spin and, for example, ask for the fraction of particles in the ensemble with spin in the $\pm x$ direction, the measurement is represented by a projection operator $P_x(\sigma) = |\sigma_x\rangle\langle\sigma_x|$, with $\sigma_x = \pm 1$. In terms of the eigenvectors of $\sigma_z$,

$$|\sigma_x = \pm 1\rangle = \frac{1}{\sqrt{2}} (|+1\rangle \pm |-1\rangle).$$

It is true (for any of the values of $\sigma_x$ and $\sigma_z$) that

$$|\langle \sigma_x | \sigma_z \rangle|^2 = \frac{1}{2}.$$

We can write this result as

$$|\langle \sigma_x | \sigma_z \rangle|^2 = \text{Tr} (P_x(\sigma') P_z(\sigma)).$$

Let us now consider a beam of spin $\frac{1}{2}$ particles with a fraction $\gamma_+$ with spin up and $\gamma_-$ with spin down in the $z$ direction ($\gamma_+ + \gamma_- = 1$). The probability to find spin up as the outcome of the experiment is

$$P_+ = |\langle \sigma'_z = +1 | \sigma_z = +1 \rangle|^2 \gamma_+ + |\langle \sigma'_z = +1 | \sigma_z = -1 \rangle|^2 \gamma_- = \gamma_+, $$

since the second term vanishes. If $\gamma_+ = \frac{1}{2}$, the result is indistinguishable from the probability to find a spin $\pm \frac{1}{2}$ in the $x$ direction in a beam of particles with definite spin in the $z$ direction.

We can write the result of the second example as

$$P_+ = \gamma_+ \text{Tr} (P'(\sigma'_z = +1) P(\sigma_z = +1)) + \gamma_- \text{Tr} (P'(\sigma'_z = +1) P(\sigma_z = -1))$$

$$= \text{Tr} (\rho P(\sigma'_z = +1))$$

for

$$\rho = \gamma_+ P'(\sigma_z = +1) + \gamma_- P'(\sigma_z = -1).$$

The operator $\rho$ is called the density operator, representing a state consisting of a mixture of components with spin up and spin down in the ensemble of possibilities. We see that, with a slight generalization of the procedure used above with $\rho_z \rightarrow \rho_0$, no matter what direction 0 we test in the experiment, the outcome $P_0$ (a linear combination of $\gamma_+, \gamma_-$ with coefficients less than unity) can never reach unity if $\gamma_+$ or $\gamma_-$ is not unity. In the first example, where we have a beam with definite
1.1 The density operator and probability

$\sigma_z$, the state is represented by a vector, and the measurement of the spin in the $z$-direction can yield probability one. For a general choice of $\gamma_{\pm}$, there is no vector that can represent the state. In the first case the state is called pure, and it can be represented by a projection into a one-dimensional subspace (in the previous example, $P_{\sigma_z} = |\sigma_z\rangle\langle\sigma_z|$. This is equivalent to specifying the vector, up to a phase, corresponding to the one-dimensional subspace. In the second case, it is called mixed and does not correspond to a vector in the Hilbert space.

It is clear from the discussion of these examples that the a priori probabilities $\gamma_{\pm}$ are essentially classical, reflecting the composition of the beam that was prepared in the macroscopic laboratory.

Although a density operator $\rho$ of the type that we have defined in this example appears to be a somewhat artificial construction, it is actually a fundamental structure in quantum statistical mechanics (Dirac, 1958). It enables one to study a complex system in the framework of an ensemble and in fact occurs on the most fundamental level of the axioms of the quantum theory.

It was shown by Birkhoff and von Neumann (1936) that both quantum mechanics and classical mechanics can be formulated as the description of a set of questions for which the answer, as a result of experiment, is “yes” or “no.” Such a set, which includes the empty set $\phi$ (questions that are absurd, e.g. the statement that the system does not exist) and the trivial set $I$ (the set of all sets, e.g. the statement that the system exists), and is closed with respect to intersections and unions, is called a lattice. A lattice that satisfies the distributive law

$$a \cap (b \cup c) = (a \cap b) \cup (a \cap c),$$

where $\cup$ represents the union and $\cap$ the intersection, is called Boolean. These operations have the physical meaning of “or” (the symbol $\cup$), in which one or the other of the propositions is true, and “and” (the symbol $\cap$), for which both must be true for the answer of the compound measurement to be “yes.” An example of such a lattice may be constructed in terms of two-dimensional closed regions on a piece of paper. This is discussed again in the appendix to this chapter.

Both classical and quantum theories may be associated with lattices in terms, respectively, of the occupancy of cells in phase space or states in the subspaces of the Hilbert space. The questions $a$ correspond, in the first case, to the phase space cells (with answer corresponding to occupancy) and in the second to the projection operators $P_\alpha$ associated with a subspace $M_\alpha$, with the answer corresponding to the values $\pm 1$ which a projection operator can have. These values correspond to evaluating the projection operator on vectors which lie within or outside the subspace.

Birkhoff and von Neumann asserted that the fundamental difference between classical and quantum mechanics is that the lattices corresponding to classical
mechanics are Boolean, and those corresponding to quantum mechanics are not. The non-Boolean structure of the quantum lattice is associated with the lack of commutativity of the projection operators associated with different subspaces:

$$a \cap (b \cup c) \neq (a \cap b) \cup (a \cap c).$$  (1.1)

This is a fundamental difference between classical and quantum statistics.

Let us illustrate this point by a simple example, again using the spin $\frac{1}{2}$ system. Each of the Pauli spin matrices has eigenvalues $\pm 1$ and is therefore associated with a set of projection operators of the form

$$P_i = \frac{1}{2} (1 \pm \sigma_i)$$

for $i = x, y, z$. Let us consider three closed linear subspaces associated with the projections into the subspaces with the $\sigma_i$ positive, i.e. with the $P_i$ defined as above with positive signs. We call these subspaces $M_x, M_y, M_z$; they correspond to propositions which are not compatible, i.e. the corresponding projection operators do not commute. We shall show explicitly, for this simple example, that

$$M_z \cap (M_x \cup M_y) \neq (M_z \cap M_x) \cup (M_z \cap M_y),$$

that is, this set of propositions is not Boolean. The construction is interesting in that it illustrates the special structure of the topology of Hilbert spaces as well as the notion of the non-Boolean lattice.

We start by constructing the union of the manifolds $M_x$ and $M_y$ by their joint linear span. Taking the standard definition of the Pauli matrices,

$$\sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix},$$

the projection operators into the subspaces with positive eigenvalues are

$$P_x = \frac{1}{2} (1 + \sigma_x) = \frac{1}{2} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix},$$

$$P_y = \frac{1}{2} (1 + \sigma_y) = \frac{1}{2} \begin{pmatrix} 1 & -i \\ i & 1 \end{pmatrix},$$

$$P_z = \frac{1}{2} (1 + \sigma_z) = \frac{1}{2} \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}.$$

The corresponding eigenvectors are given by projecting a generic vector $v$ into the respective subspaces. For

$$v = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix},$$
1.1 The density operator and probability

using the result just given,

\[ P_x v = \frac{1}{2} (v_1 + v_2) \begin{pmatrix} 1 \\ 1 \end{pmatrix}, \]

so that \( M_x \) is represented by the linear span of the normalized eigenvector:

\[ v_x = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \end{pmatrix}. \]

Similarly,

\[ P_y v = \frac{1}{2} (v_1 - i v_2) \begin{pmatrix} 1 \\ i \end{pmatrix}, \]

so that the corresponding (normalized) eigenvector is

\[ v_y = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ i \end{pmatrix}. \]

Finally,

\[ P_z v = v_1 \begin{pmatrix} 1 \\ 0 \end{pmatrix}, \]

so the corresponding eigenvector is

\[ v_z = \begin{pmatrix} 1 \\ 0 \end{pmatrix}. \]

The union of the subspaces \( M_x \) and \( M_y \) is the closed linear span of vectors in both subspaces. By taking the combination \( v_x + i v_y \), it is easy to see that the vector \( v_z \) (and hence the subspace \( M_z \)) is contained in \( M_x \cup M_y \). To construct the distributed operation

\[ (M_z \cap M_x) \cup (M_z \cap M_y), \]

we must use the construction for which the projection operator corresponding to the intersection of two noncompatible subspaces is generated by an alternating succession of projections into the two subspaces (Jauch, 1968). The products \( P_z P_x \) and \( P_z P_y \) are, it so happens, idempotents up to coefficients less than one, i.e.

\[ P_z P_x = \frac{1}{2} \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}, \]

\[ (P_z P_x)^2 = \frac{1}{4} \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}. \]
Foundations of quantum statistical mechanics

and

\[
P_z P_y = \frac{1}{2} \begin{pmatrix} 1 & -i \\ 0 & 0 \end{pmatrix}
\]

\[
(P_z P_y)^2 = \frac{1}{4} \begin{pmatrix} 1 & -i \\ 0 & 0 \end{pmatrix},
\]

which implies that both \((P_z P_x)^n\) and \((P_z P_y)^n\) go to zero as \(n \to \infty\). Therefore,

\[M_z \cap M_x = M_z \cap M_y = 0.\]

Clearly,

\[M_z \cap (M_x \cup M_y) \neq (M_z \cap M_x) \cup (M_z \cap M_y).\]

Although \(P_z P_x\) and \(P_z P_y\) are not zero (the two corresponding vectors are not orthogonal), the closed subspace that is common is empty. One can think of this geometrically in terms of two lines that have some projection on the other, but the intersection of the two lines is just a point of zero measure. Physically, this implies that we cannot have a definite statement of the joint values of \(\sigma_z\) and \(\sigma_x\) or \(\sigma_y\). The noncommutativity of the associated projections is essential; if they were commutative, the product of projections would be a projection, and the products would not converge to zero. It is clear from this example that compatible subspaces would satisfy Boolean distributivity.

We shall later discuss the Wigner function, which appears to provide joint distributions over noncommutative variables such as \(q\) and \(p\); however, these functions are not probabilities, since, although they are the coefficients of what might be called the Weyl basis for the operator algebra of the quantum theory which appear in expectation values, they are not positive (Wigner, 1936).

1.2 The Gleason theorem and consequences

The axioms of quantum mechanics are implicitly developed in the fundamental work of Dirac (1958). Let us focus here on probability. Given \(P_i (i = 1, \ldots)\), a sequence of projections \(P_i P_k = 0\) for \(i \neq k\), then the probability measure \(w\)

\[w : P \to [0, 1]\]
1.2 The Gleason theorem and consequences

satisfies

(a) \[ \bigcup_i w(P_i) = w\left(\sum_i P_i\right) \] (1.2)

(b) \[ w(\phi) = 0, \quad w(I) = 1 \]

(\(\phi\) is the zero projection)

(c) \[ w(P) = w(F) = 1 \rightarrow w(P \cap F) = 1 \]

Piron (1976) added another axiom, namely that partially ordered (by inclusion) sets of the non-Boolean lattice of the quantum theory form Boolean sublattices, and with this he was able to show a converse result, i.e. that such partially ordered lattices can be embedded in a Hilbert space (or a family of Hilbert spaces if there are superselection rules), thus inducing the full structure of the quantum theory.

Along with the sets of “yes-no” questions that form the basic elements of the quantum lattice, one may assume a function \( w(a) \) with values between zero and unity, with the interpretation of a probability measure, which has the so-called sigma additivity property

\[ w(a \cup b) = w(a) + w(b) \] (1.3)

when \( a \) and \( b \) have no intersection, i.e. \( a \cap b = \phi \). This idea is consistent with the notion of probability for the “yes” answer for \( a \) and \( b \). Gleason (1957) showed that for any Hilbert space of three or more real dimensions, there is a density operator, self-adjoint and positive, \( \rho \), such that

\[ w(a) = \text{Tr} \rho P_a, \] (1.4)

where \( P_a \) is the projection operator into a subspace corresponding to the question \( a \). This existence theorem is one of the most powerful and important theorems in the foundations of the statistical quantum theory. The function \( w(a) \) is called a state, a notion completely consistent with Dirac’s definition of a state in the quantum theory, i.e. for any \( a \), this function provides the probability of its truth and therefore corresponds to maximum knowledge.

The original proof of Gleason is rather long and involved, but Piron has given a simple and elegant proof, which is given here in an appendix to this chapter for the mature student.

The density operator (often called “density matrix”) has the properties

\[ \text{Tr} \rho = 1 \] (1.5)

\[ \text{Tr} \rho^2 \leq 1. \]

The first follows from the fact that the sum over all disjoint \( a \) of \( w(a) \) is the total probability measure on the set of all questions (and the sum over all disjoint \( P_a \) is
the unit operator). The second follows from the first; all eigenvalues of $\rho$ are real and positive with values less than or equal to unity. With these properties, one can prove that the spectrum of $\rho$ must be completely discrete.

Mackey (1963) has given a converse theorem. If the function $w(a)$ can reach the value unity on a one-dimensional subspace of the Hilbert space, the corresponding density operator is just a projection into this one-dimensional subspace and can be put into correspondence (up to a phase) with the vector of the Hilbert space generating this one-dimensional subspace. Such a state is called pure. A state which cannot reach the value of unity on any one-dimensional subspace is called mixed.

The proof is very simple. Let $P_0$ be the projection onto a one-dimensional subspace generated by the vector $\phi_0$, and let us use the representation, taking into account the discrete spectrum of $\rho$,

$$\rho = \sum_i \gamma_i \langle \psi_i | \psi_i \rangle.$$

(1.6)

Here we use the Dirac ket $| \psi_i \rangle$ to signify an element of the Hilbert space. Then if $\text{Tr} \rho P_0 = 1$, it follows that $\text{Tr} \rho (1 - P_0) = 0$, or

$$\text{Tr} \sum_i \gamma_i \langle \psi_i | (1 - P_0) | \psi_i \rangle = \text{Tr} \sum_i \gamma_i \| (1 - P_0) | \psi_i \rangle \|^2 = 0,$$

where $\| \chi \|^2$ is defined as $\langle \chi | \chi \rangle$, the norm of the vector $| \chi \rangle$. Since the $\gamma_i$ are positive, this implies that

$$(1 - P_0) | \psi_i \rangle = 0$$

for all of the $| \psi_i \rangle$, i.e.,

$$| \psi_i \rangle = \lambda_i | \phi_0 \rangle$$

for all $i$. Substituting into Eq. (1.6), we see that in this case we must have

$$\rho = \sum_i \gamma_i | \lambda_i |^2 | \phi_0 \rangle \langle \phi_0 |.$$

Furthermore, if the $| \psi_i \rangle$ and $| \phi_0 \rangle$ are normalized, $| \lambda_i |^2 = 1$. Then, by Eq. (1.5) and Eq. (1.6) (for the $| \psi_i \rangle$ orthogonal), one sees that the sum of the $\gamma_i$ is unity; hence

$$\rho = | \phi_0 \rangle \langle \phi_0 |,$$

which is the projection operator into the subspace generated by $| \phi_0 \rangle$. This theorem therefore identifies the pure states with vectors of the Hilbert space, and it is for this reason that one often calls the vectors of the Hilbert space “states.” Every vector in the Hilbert space corresponds to a pure state.
1.3 Calculation of averages of observables

If $w_1$ and $w_2$ are two different states, then

$$w = \lambda_1 w_1 + \lambda_2 w_2$$

with $\lambda_1 + \lambda_2 = 1$ and with $\lambda_1, \lambda_2$ positive also is a state; the set of states form a convex set (Jauch, 1968). Such a state is called a mixture. A state which cannot be represented in terms of two others is called pure; the pure states are the extremal subset of a convex set. These definitions are, of course, consistent with Mackey’s result.

1.3 Calculation of averages of observables

Let us now consider an observable represented by a self-adjoint operator $A$ on the Hilbert space with a spectrum of discrete eigenvalues $a_k$. Such an operator can be represented as a sum over projections into its eigenstates, i.e.

$$A = \sum_k a_k P_k, \quad (1.7)$$

where, if $P_k = |\phi_k\rangle \langle \phi_k|$ and the $|\phi_k\rangle$ form a normalized orthogonal set, we clearly have

$$A |\phi_k\rangle = a_k |\phi_k\rangle.$$ 

The expectation of this operator in some pure state represented by $|\psi_i\rangle$ is then

$$\langle \psi_i | A | \psi_i \rangle = \sum_k a_k \langle \psi_i | P_k | \psi_i \rangle = \sum_k a_k ||\langle \psi_i | \phi_k \rangle||^2, \quad (1.8)$$

with the usual quantum interpretation that $||\langle \psi_i | \phi_k \rangle||^2$ is the quantum mechanical probability that a system in the state described by $|\phi_k\rangle$ is found in the state $|\psi_i\rangle$. The weighting of the eigenvalues of $A$ by this probability then gives the expected value of this observable in the state described by $|\psi_i\rangle$. Suppose now that we prepare a system which contains subsystems in the states $|\psi_i\rangle$ according to the $a priori$ probability distribution $\gamma_i$. This can be arranged by preparing a system with the number of subsystems in each state $|\psi_i\rangle$ proportional to the $\gamma_i$. This is an ensemble. We emphasize here that this step, as in our previous example, is entirely classical. We build an ensemble of subsystems with $a priori$ probabilities based on their frequency of occurrence, a completely classical notion of probability, i.e. the frequency interpretation.

The overall expectation of the value of the observable $A$ is then given by the sum over all of the expected values in each of the quantum states, with coefficients
equal to the classical probabilities of the occurrence of each quantum state in the
ensemble, i.e.

\[ \langle A \rangle = \sum_i \gamma_i \langle \psi_i | A | \psi_i \rangle . \]

This result is obtained directly by computing

\[ \langle A \rangle = \text{Tr} \rho A, \]

where

\[ \rho = \sum_i \gamma_i |\psi_i \rangle \langle \psi_i | . \]

Viewing this in a slightly different way, we see that

\[ \langle A \rangle = \sum_k a_k \text{Tr} (\rho P_k) , \]

where

\[ \text{Tr} (\rho P_k) = \sum_i \gamma_i \langle \psi_i | P_k | \psi_i \rangle \]

\[ = \sum_i \gamma_i |\langle \psi_i | \phi_k \rangle|^2 \]

is the probability of finding the system in the subspace associated with \( P_k \). This
probability is composed of two types of expectation: the quantum probability to
find the \( P_k \) in each state \( \psi_i \), and the classical probability for the occurrence of the
state \( \psi_i \) (determined by the relative number of subsystems in that state).

The results that we have given can easily be extended to the most general case
of an observable with both discrete and continuous spectra without change in the
formal structure, although as we shall see later, there are special technical aspects
that arise in the continuous case (for example, in scattering theory). To see this,
we use the spectral representation theory of von Neumann. It was shown by von
Neumann (1955) that every self-adjoint operator \( A \), corresponding to a physical
observable, has a spectral representation of the form

\[ A = \int a \, dE (a) , \]

where \( a \) takes on a continuous set of values (the real line), and the self-adjoint set
of operators \( E (a) \) is called a “spectral family.” It satisfies the property

\[ E (a) E (b) = E (\min (a, b)) , \]